Statistical and Intelligent Methods of Medical Data Processing

G.R. Shakhmametova, N.I. Yusupova, V.V. Mironov
Computer Science & Robotics Department
Ufa State Aviation Technical University
Ufa, Russia

R.Kh. Zulkarneev
Faculty of General Medicine
Bashkir State Medical University
Ufa, Russia

***Abstract*—The new approach to the medical, in particular, the toxicological data analysis is considered. For the data processing multilevel system realization the three-stage technique for data analysis is offered what allows to reach the comprehension about the data structure, to extract patterns, to get new, unknown knowledge, and also to increase the data analysis process efficiency. The results of the research are discussed.**

Keywords—medical data; data processing; non-parametric methods; data mining

I. INTRODUCTION

Contemporary application of data computer analysis methods is widely used in all spheres of the human activity. Decision-making in many areas, including medicine and health care, is based on the data analysis. By means of the data analysis it is possible to confirm or disprove the assumption, to significantly prove efficiency of treatment, to increase efficiency of diagnostics, to find the patterns, and to forecast the results of treatment [1].

The most important part of IT using in medicine is processing of data intended for solving diagnostics problems and diseases treatment, especially when managing large volume of the input information or realization of complicated algorithms of data processing is difficult for the practical doctor or user [2]. Therefore, using information technologies for development and creation of the data analysis techniques necessary for decision-making is crucial. The importance of this task increases with avalanche-like accumulation of information today due to improvement of collecting and storage information technologies. Introduction of information technologies in medical practice allows to change radically a situation though there is a lot of problems which need to be solved by joint efforts of information scientists and physicians.

In this article, medical data analysis in the field of toxicology is considered and examples of the analysis of the toxicological data collected in the Republic of Bashkortostan (RB) for 2015-2016 with interpretation are given. The analysis of these data allows to make reasonable decisions in the field of treatment and prevention of toxicological diseases that is important not only from the medical, but also the social point of view.

II. STATE OF ART

Today one of the most popular medical data in the field of toxicology analysis methods is the elementary visual analysis with diagrams and charts use. In general for the medical data study the statistical and intelligent analysis techniques are used. In [3] the medical data analysis methodology is considered, the research is aimed at doctors-clinical physicians activities efficiency increasing. The preliminary analysis was carried out with the charts of dispersion and density use, the applied analysis methods - correlation analysis, logistic regression, the accidental trees nonparametric qualifier. In [4] the various selection methods which may be used in medical researches with various scenarios and problems are considered. The main applied methods - sampling and randomization. A large number of researches are devoted to Data Mining applications for the hidden patterns recognition in the medical data analysis tasks. For example, in [5] the Data Mining use for obesity disease detection at children is discussed. The cluster analysis use for definition of children's groups with similar results after the executed treatment is considered. The methods applied in the research - dispersion analysis, cluster analysis (K-averages method), the limited search algorithm. In [6] the analysis of data on acute oral poisonings based on the REACH data is considered, the main research direction – the separate chemicals impact definition on an organism. The data analysis is carried out by following methods - training at examples, neural networks, k-nearest neighbors. In [7] the research of acute exogenous poisonings in Altai Region during 1997-2013 is given. The research allows to assess a toxicological situation in the region and includes the systematization of acute poisonings in a section of gender and age and social groups of the population. Researches were carried out with data visual analysis methods use in MS Excel. In [8] the research of possible risks at patients with acute alcohol poisoning is conducted, the descriptive statistics and the visual analysis methods are used.

All data analysis researches considered by authors use limited quantity of methods – or only the visual analysis, or only the statistical methods, or only the Data mining methods. Thus, today there is no complex technique for the medical data analysis allowing to process data most fully and to get the maximum quantity of new knowledge and hidden patterns.

Medical data, from the point of view of the analyst, have the next features:

The research is supported by Russian Foundation for Basic Research, grants № 15-07-01565, № 16-07-00239, № 16-07-00773.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

1) the data are retrospective;

2) as a rule, they are diverse and have quantitative and quality indicators;

3) they are semi structured;

4) as a rule, such data do not submit to normal distribution.

In this regard, there are difficulties with processing and analysis of such data because of heterogeneity of the data that demands the application of various methods and approaches for data analysis.

Several methods are used today for processing of medical data [9-13]: descriptive and inductive statistics; correlation, regression, multiple-factor analysis; method of artificial neural networks, and others. Comparative characteristics of these methods are provided in Tab. 1.

TABLE I. COMPARATIVE CHARACTERISTICS OF THE DATA PROCESSING METHODS IN MEDICINE

Method	Goal	Advantages	Shortcomings
descriptive statistics	Systematization and description of observation data	Effective and rather easy way of data consideration and description; convenient way of information representation [14]	The restrictions connected with the size of selection and the used method [15]
inductive statistics	Research of distinctions between indicators in various samplings for their statistical importance or insignificance detection [16]	Simplicity of method application	Low level of reliability; considerable errors in the case of small size samplings
correlation analysis	Detection of the fact of existence and level of communication between two and more variables for predicting possible value of one of them if another is known [17]	Possibility of creating new rules for interaction of functions and also an assessment of functions interaction [18]	The results may be used only in the immediate research field or in one close to it
regression analysis	Detection of the fact of dependence between independent variable and one or several dependent variables [19]	It allows to submit data of a response in summary form, to compare various but connected data sets and to Analyze potential relations "the reason - the investigation" [19, 20]	Existence of unaccounted variables, errors of measurements and other sources of not explained Variations of a response can complicate simulation; including unnecessary variables can hide influence of independent variables and reduce a forecast accuracy [21]
multiple-factor analysis	Detection of latent variables or the factors causing multiple correlative communica-tions [13]	Possibility of smaller number of data using and, therefore, leading to more expedient model generation; white-box method, i.e., it is completely open and clear [13, 22]	Difficulties in factors sampling causing subjectivity of result interpretations; the multiple-factor analysis is a complicated procedure [13, 22]
neural networks	Generalization and allocation of hidden dependences between the input and output data	No need of knowledge formalization; orientation to parallel processing; possibility of multidimensional data and knowledge processing without increase in labor input [23]	Difficulties in explanation of neural network work results; impossibility to guarantee repeatability and uniqueness of obtaining results [24]

Each of these methods has its advantages and shortcomings, and many of them have restrictions on character of the analyzed data. The problem is that a single method may only solve a narrow task of the data analysis which is not enough for decision-making. Authors offer a complex technique of the analysis of medical data in the field of toxicology including methods of mathematical statistics as well as methods of data mining, allowing to carry out the comprehensive analysis of the data and to benefit from the largest possible amount of knowledge, interrelations and patterns. The research novelty consists in the new approach to the toxicology data analysis which represents the complex three-stage analysis with visual, statistical and Data Mining methods use and allows to study the data comprehensively.

III. DATA ANALYSIS STAGES

The technique of medical data analysis in the field of toxicology including complex data analysis and interpretation of results, and consisting of the next main stages is suggested as follows (fig. 1):

1) Primary statistical data analysis by the means of MS Excel, visualization of raw data for understanding their quantitative and qualitative structure, making hypotheses about patterns existence in data.

2) Statistical analysis of the data using multidimensional analysis and nonparametric methods for confirmation or denial of the hypotheses made at the first stage. This stage is also the stage of the "prospecting" analysis for the purpose of making new hypotheses and assumptions.

3) Data mining assumes search of the hidden regularities, patterns in the data with use of means of Data mining, knowledge discovery, confirmation or denial of the hypotheses made at the previous stages.

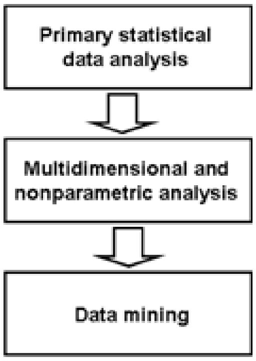

Primary statistical data analysis
Multidimensional and nonparametric analysis
Data mining

Fig. 1. Stages of medical data analysis in the field of toxicology.

For the analysis, the sampling of the toxicological data on the Republic of Bashkortostan for 2015-2016 has been taken. The volume of the data base for the analysis is 6 338 records. The structure of the data includes:

- gender;
- social group;
- address (city, region);
- city/village;
- place of poisoning;
- date of poisoning;
- diagnosis;
- MKB10 code;
- who has made the diagnosis;
- lethal outcome;
- individual/group;
- number of victims;
- purpose;
- place of obtaining poison;
- health facility;
- other.

The data are diverse, and only a part of the data is quantitative (numerical); the most part of data is qualitative (symbolical) which, on the one hand, complicates the statistical analysis, but on the other hand this creates prerequisites of data mining application.

A. The Results of the Primary Statistical Data Analysis

The primary research of raw data is the first stage of the analysis and is carried out for the detection of the most general regularities and tendencies, character and properties of the analyzed data, and laws of the analyzed data distribution [25]. The results of the initial prospecting analysis are not used for making decisions, their –purpose is to help in the development of the best strategy of the profound analysis, hypotheses making, specification of these or those mathematical methods and models features application. The prospecting analysis helps to concisely describe the structure of data in a visual form, and then to research it in more detail by means of statistical analysis and data mining. The purpose of this stage is to visualize data and to collect the maximum quantity of hypotheses for possible interrelations and regularities in data.

In Fig. 2-4, examples of first stage realization of the data analysis are presented. On the basis of visualization of the data, it is possible to make assumptions of quantitative and qualitative structure of data, and also about interaction of various factors.

The most frequent reason of acute poisonings in RB is alcohol (47,8%), on the second place there are drugs (37,88%), further – narcotic substances (5,99%), carbon monoxide (5,43%), mushrooms (2,15%) and snake bites (0,74%) (Fig. 2).

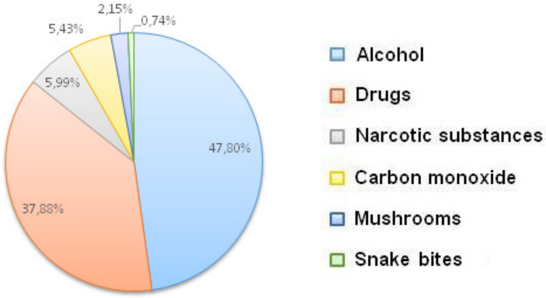


Fig. 2. Main reasons of acute poisonings.

From the known substances which caused acute poisonings with a lethal outcome in 2016 in RB, on the first place is alcohol (28,9%), on the second is carbon monoxide (11,1%); further there are narcotic substances (7,2%), drugs (1,8%) corroding substances (1,6%), and pesticides (0,2%), while the greatest sector, nearly a half, is made of other unspecified substances (49,2%) (Fig.3).

From the chart in Fig. 4, it is possible to see that the peak of acute poisonings for males is at the age of 46-60 years, for females – at the age over 75; however, that is non-evident knowledge and demands further research.

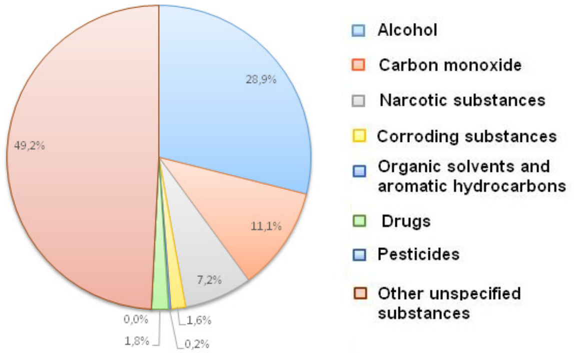


Fig. 3. Structure of the poisons which have caused acute poisonings.

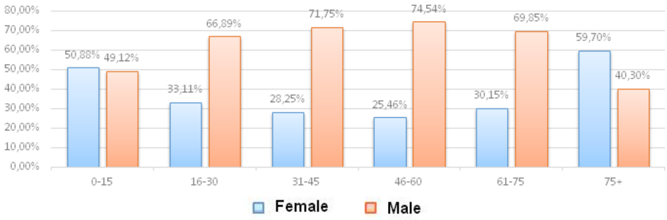


Fig. 4. Structure of acute poisonings depending on age for male and female.

B. *The Results of the Data Statistical Analysis*

The next stage of analysis is deeper data studying by means of the statistical analysis methods.

According to statistical principles used in the basis, the methods are subdivided in [26]:

- parametric - applied mainly to the analysis of normally distributed quantitative signs);
- nonparametric - applied to the analysis of quantitative signs irrespective of their distribution and to the analysis of qualitative signs).

Nonparametric methods are developed for those situations when the researcher knows nothing about parameters of the analyzed data. Owing to features of the medical data which are listed above it is more effective to apply methods of the nonparametric analysis to their processing [27].

The examples of data processing results with use of nonparametric methods of the analysis executed in version 13.2 STATISTICA package are given below.

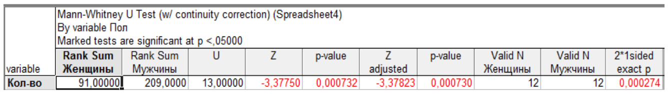
Mann-Whitney's criterion

Mann-Whitney's U-criterion is statistical criterion which is used at an assessment of distinctions between two independent samplings on the level of any sign measured quantitatively [28]. The statistics of criterion looks as follows:

$$U = W - \frac{1}{2} m(m+1) = \sum_{i=1}^{n} \sum_{j=1}^{m} \delta_{ij} \qquad (1)$$

where W – Vilkokson's statistics, intended for check of the same hypothesis H0.

For Mann-Whitney's test, the Statistics Nonparametrics module and the procedure of this module, i.e., *Comparing two independent samples module (groups)* (Fig. 5) have been used.



Mann-Whitney U Test (w/ continuity correction) (Spreadsheet4)
By variable Пол
Marked tests are significant at p <,05000

variable	Rank Sum Женщины	Rank Sum Мужчины	U	Z	p-value	Z adjusted	p-value	Valid N Женщины	Valid N Мужчины	2*1sided exact p
Кол-во	91,00000	209,0000	13,00000	-3,37750	0,000732	-3,37823	0,000730	12	12	0,000274

Fig. 5. Mann-Whitney's criterion.

The following designations are accepted:

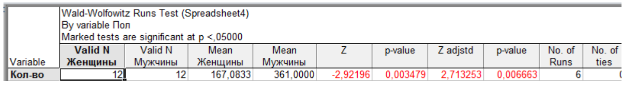
- Rank Sum T_i – sum of T_i sampling ranks;
- U – Mann-Whitney statistics for small samplings;
- p-level – probability of acceptance of the hypothesis H_0;
- p-level – corrected probability of acceptance of the hypothesis of H_0;
- Rank Sum T_j – sum of T_j sampling ranks ;
- Z – normal approximation of Mann - Whitney statistics for big samplings;
- Z adjusted – corrected normal approximation of Mann - Whitney statistics;
- Valid N – sampling volume;
- 2*1 sided exact p – probability of p are equal 1 minus cumulative unilateral probability of the corresponding Mann - Whitney statistics.

The critical value of criterion is found according to the special table. Let the significance value be equal to 0.05. H0 hypothesis of insignificance of distinctions between points of two classes is accepted if $u_{kp} < u_{emp}$. Otherwise, H_0 is rejected and distinction is defined as essential. Test results show the significant distinction between quantity of poisonings for males and females (p =0.0007 < 0.05).

Wald-Wolfowitz's criterion

Wald-Wolfowitz's criterion is the nonparametric criterion also known as the test of series. It is based on the analysis of regularities of the sequence of objects distribution of two untied samplings [29]. It can be applied in the analysis of quantitative, rank and alternative variations. Owing to ist originality, has limited application though in some cases may be more effective than other methods. The statistics of criterion of Wald-Wolfowitz suggested in [30] is based on coefficient of serial correlation and looks like this (Fig. 6):

$$R_1 = \sum_{i=1}^{n-1} x_i x_{i+1} + x_n x_1 \qquad (2)$$



Wald-Wolfowitz Runs Test (Spreadsheet4)
By variable Пол
Marked tests are significant at p <,05000

Variable	Valid N Женщины	Valid N Мужчины	Mean Женщины	Mean Мужчины	Z	p-value	Z adjstd	p-value	No. of Runs	No. of ties
Кол-во	12	12	167,0833	361,0000	-2,92196	0,003479	2,713253	0,006663	6	(

Fig. 6. Wald-Wolfowitz's criterion.

Test results also show the high importance of distinction between poisoning indicators for males and females.

At the second analysis stage of data, deep interrelations between data which can be used for decision-making are detected. At the same time, both stages are intended for the prospecting analysis, the best understanding of data and hypotheses making which are preliminary steps for data mining.

C. *The Results of Data Mining Stage*

The technology of data mining allows to discover such patterns among large volumes of data which cannot be found by statistical ways of data processing, but are objective and practically useful. Using these methods. the researcher can observe five main patterns in data [31], namely:

- association – several events are connected with each other;
- sequence – a chain of the events connected in time;
- classification – reference of an object to one of classes with the known characteristics;
- clustering – allocation of uniform groups of objects;
- temporary templates - dynamics of behavior of target indicators.

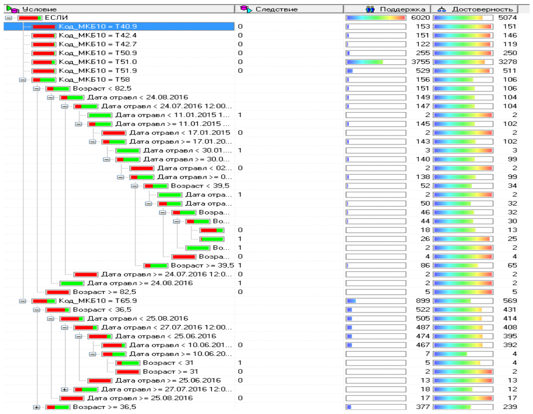
All listed patterns are applicable to medical data.

In Fig. 7-8, examples of results of the data processing with use of decision trees executed in Deductor Studio Academic 5.3.0.88 are presented. In Fig. 7, the decision tree for definition of poisoning outcomes depending on the MKB10 code, age and date of poisoning is presented.

The rules created for the decision tree presented in Fig. 7 are shown in Fig. 8.

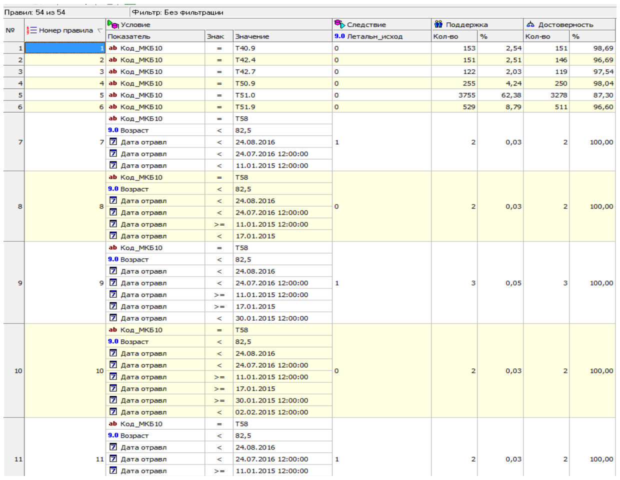
In total, 54 rules have been created in this example. I.e., the rule № 9 is follows:

IF MKB10_code = T58 AND Age < 82,5 years AND Date_of_poisoning < 24.08.2016 AND Date_of_poisoning >= 17.01.2017 THEN Lethal_outcome



Условие	Следствие	Поддержка	Достоверность
ЕСЛИ		6020	5074
Код_МКБ10 = T40.9	0	153	151
Код_МКБ10 = T42.4	0	151	146
Код_МКБ10 = T42.7	0	122	119
Код_МКБ10 = T50.9	0	255	250
Код_МКБ10 = T51.0	0	3755	3278
Код_МКБ10 = T51.9	0	529	511
Код_МКБ10 = T58		156	106
Возраст < 82,5		151	106
Дата отравл < 24.08.2016		149	104
Дата отравл < 24.07.2016 12:00...		147	104
Дата отравл < 11.01.2015 1...	1	2	2
Дата отравл >= 11.01.2015 ...		145	102
Дата отравл < 17.01.2015	0	2	2
Дата отравл >= 17.01.20...		143	102
Дата отравл < 30.01....	1	3	3
Дата отравл >= 30.0...		140	99
Дата отравл < 02...	0	2	2
Дата отравл >= 0...		138	99
Возраст < 39,5		52	34
Дата отра...	1	2	2
Дата отра...		50	32
Возра...		46	32
Во...		44	30
	0	18	13
	1	26	25
Во...	1	2	2
Возра...	0	4	4
Возраст >= 39,5	1	86	65
Дата отравл >= 24.07.2016 12:0...	0	2	2
Дата отравл >= 24.08.2016	1	2	2
Возраст >= 82,5	0	5	5
Код_МКБ10 = T65.9		899	569
Возраст < 36,5		522	431
Дата отравл < 25.08.2016		505	414
Дата отравл < 27.07.2016 12:00...		487	408
Дата отравл < 25.06.2016		474	395
Дата отравл < 10.06.201...	0	467	392
Дата отравл >= 10.06.20...		7	4
Возраст < 31	1	5	4
Возраст >= 31	0	2	2
Дата отравл >= 25.06.2016	0	13	13
Дата отравл >= 27.07.2016 12:0...		18	12
Дата отравл >= 25.08.2016	0	17	17
Возраст >= 36,5		377	239

Fig. 7. The decision tree for definition of poisoning outcomes.



Правил: 54 из 54 | Фильтр: Без фильтрации

№	Номер правила	Условие			Следствие	Поддержка		Достоверность	
		Показатель	Знак	Значение	9.0 Летальн_исход	Кол-во	%	Кол-во	%
1	1	ab Код_МКБ10	=	T40.9	0	153	2,54	151	98,69
2	2	ab Код_МКБ10	=	T42.4	0	151	2,51	146	96,69
3	3	ab Код_МКБ10	=	T42.7	0	122	2,03	119	97,54
4	4	ab Код_МКБ10	=	T50.9	0	255	4,24	250	98,04
5	5	ab Код_МКБ10	=	T51.0	0	3755	62,38	3278	87,30
6	6	ab Код_МКБ10	=	T51.9	0	529	8,79	511	96,60
7	7	ab Код_МКБ10	=	T58	1	2	0,03	2	100,00
		9.0 Возраст	<	82,5					
		Дата отравл	<	24.08.2016					
		Дата отравл	<	24.07.2016 12:00:00					
		Дата отравл	<	11.01.2015 12:00:00					
8	8	ab Код_МКБ10	=	T58	0	2	0,03	2	100,00
		9.0 Возраст	<	82,5					
		Дата отравл	<	24.08.2016					
		Дата отравл	<	24.07.2016 12:00:00					
		Дата отравл	>=	11.01.2015 12:00:00					
		Дата отравл	<	17.01.2015					
9	9	ab Код_МКБ10	=	T58	1	3	0,05	3	100,00
		9.0 Возраст	<	82,5					
		Дата отравл	<	24.08.2016					
		Дата отравл	<	24.07.2016 12:00:00					
		Дата отравл	>=	11.01.2015 12:00:00					
		Дата отравл	>=	17.01.2015					
		Дата отравл	<	30.01.2015 12:00:00					
10	10	ab Код_МКБ10	=	T58	0	2	0,03	2	100,00
		9.0 Возраст	<	82,5					
		Дата отравл	<	24.08.2016					
		Дата отравл	<	24.07.2016 12:00:00					
		Дата отравл	>=	11.01.2015 12:00:00					
		Дата отравл	>=	17.01.2015					
		Дата отравл	>=	30.01.2015 12:00:00					
		Дата отравл	<	02.02.2015 12:00:00					
11	11	ab Код_МКБ10	=	T58	1	2	0,03	2	100,00
		9.0 Возраст	<	82,5					
		Дата отравл	<	24.08.2016					
		Дата отравл	<	24.07.2016 12:00:00					
		Дата отравл	>=	11.01.2015 12:00:00					

Fig. 8. The rules created for the decision tree.

The rules obtained are already the new, unevident knowledge taken from data and suitable for use at decision-making.

IV. CONCLUSION

The suggested technique of the medical data analysis in the field of toxicology is geared toward supporting the process of decision making and includes three main stages, such as primary analysis of data, application of methods of nonparametric statistics and data mining. For the analysis, a sampling of the toxicological data collected in the Republic of Bashkortostan during 2015-2016 was taken. At the first stage, the visualization of the data for the best understanding of their structure (in particular, the main reasons for acute poisonings, structure of poisons and many others things are defined), making the assumptions of existing interrelations in data, receiving prior knowledge (for example, and age peaks of acute poisonings for males and females) is carried out. At the second stage, the hypothesis and assumptions are checked by the means of statistical methods. As medical data are analyzed, the nonparametric methods allowing processing the data which are not submitting to normal distribution are applied. In the reviewed examples by the means of Mann-Whitney's and Wald-Wolfowitz's criteria existence of significant distinctions between quantity of poisonings and indicators of poisonings depending on gender is proved. Data mining allows to reveal hidden patterns in the analyzed data and to find new, earlier unknown knowledge. Formation of a decision tree for definition of poisonings outcome is shown, and by means of technology of data mining, 54 rules for the specified decision tree are formulated. The authors intend to continue further research in the field of expansion in data mining application for extraction of implicit regularities and patterns from medical toxicological data.

REFERENCES

[1] B.A. Kobrinsky, T.V. Zarubin, Medical Informatics. Moscow: Prod. Akademiya center, 2009.

[2] I.P. Korolyuk, Medical informatics, 2nd prod. Samara: GBOU VPO “SamGMU”, 2012.

[3] Tsanas, M.A. Little, P.E. McSharry, “A methodology for the analysis of medical data,” in Handbook of Systems and Complexity in Health, Springer, New York, 2013, pp. 113-125.

[4] Karthik Suresh, Sanjeev V.Thomas, Geetha Suresh, “Design, data analysis and sampling techniques for clinical research,” Ann Indian Acad Neurol, 2011 Oct-Dec; 14(4), pp. 287–290.

[5] O.V. Marukhina, E.E.Mokina, E.V. Berestneva, “Using Data Mining for revealing hidden regularities in the task of analyzing medical data,” in Fundamental Research, 2015, vol. 4, pp. 107-113.

[6] Thomas Luechtefeld, Alexandra Maertens, Daniel P. Russo, Costanza Rovida, Hao Zhu and Thomas Hartung, “Analysis of Public Oral Toxicity Data from REACH Registrations 2008-2014,” Alternatives to Animal Experimentation : ALTEX , 2016, vol. 33 (2), pp. 111-122.

[7] I.P. Saldan, A.A. Ushakov, T.N. Karpova, “The analysis of the situation on chemical etiology acute poisonings in the Altai Region administrative center city Barnaul in 1997-2013,” Fundamental and application-oriented aspects of risk analysis for the population health: Materials of the All-Russian scientific and practical Internet-conference of young scientists. Perm: Book format, 2014, pp. 121-128.

[8] Joachim Gruettner, Thomas Walter, Siegfried Lang, Miriam Reichert, Stephan Haas, “Risk Assessment in Patients with Acute Alcohol Intoxication,” in Vivo, 2015. vol. 29, no. 1, pp. 123-127.

[9] S.V. Chebotaryov, “Theory and practice of the static and dynamic economic factorial analysis,” in Control systems and information technologies. Voronezh: Central Chernozem book publishing house, 2001, pp. 68-73.

[10] I.S. Enyukov, The factorial, discriminant and cluster analysis. Moscow: Finance and statistics, 1989.

[11] A.N. Krichevets, A.A. Korneev, E.I. Rasskazova, Mathematical statistics for psychologists. Moscow: Academia, 2012.

[12] A.G. Kochetov, O.V. Lyang., V.P. Masenko, Methods of statistical processing of medical data: Methodical recommendations for interns and graduate students of medical educational institutions, scientists. Moscow: RKNPK, 2012.

[13] G.N. Ovsyannikov, The factorial analysis in an available statement: studying of multiple parameter systems and processes. Moscow: Librikon, 2013.

[14] A. M. Ilyshev, O. M. Shubat, General theory of statistics. Manual. Moscow: Knorus, 2013.

[15] V. V. Yefimov, Statistical methods in quality management. Manual. Ulyanovsk: UlGTU, 2003.

[16] O. Yu. Yermolaev-Tomin, Mathematical methods in psychology, 5th prod. Moscow: Yurait, 2014.

[17] F. I. Karmanov, V. A. Ostreykovsky, Statistical techniques of processing of the experimental data. A laboratory practical work with use of a packet of MathCAD. Moscow: Contour, 2012.

[18] A. I. Kobzdar, Applied mathematics and statistics. Moscow: FIZMATLIT, 2012.

[19] A. Nasledov, SPSS 19. Professional statistic analysis of data. SPb.: St. Petersburg, 2011.

[20] N. I. Sidnyaev, Theory of planning of an experiment and analysis of statistical data. Manual, – 2nd prod. Moscow: Yurait, 2015.

[21] I. I. Yeliseyev, M. M. Yuzbashev, General theory of statistics. Moscow: Finance and statistics, 2004.

[22] E. Filatov, Methods of the determined (functional) factor analysis. LAP Lambert Academic Publishing, 2012.

[23] D. M. Eremin, I. B. Gartseev, Artificial neural networks in intellectual management systems. Moscow: MIREA, 2004.

[24] S.Khaykin, Neural Networks: A Comprehensive Foundation, 2nd prod. Moscow: Williams, 2006.

[25] E. A. Vukolov, Bases of the statistical analysis. Workshop on statistical methods and research of operations with use of STATISTICA and EXCEL packages. Moscow: Forum, 2010.

[26] P. F. Askerov, R. N. Pakhunova, A. V. Pakhunov, General and applied statistics. Moscow: Infra-M, 2014.

[27] V. A. Akulov, “The nonparametric analysis in health protection tasks,” News of the Samara scientific center of the Russian Academy of Sciences, 2013, vol. 1, pp. 2-3.

[28] E. R. Goryainova, A. R. Punks, E. N. Platonov, Applied methods of the analysis of statistical data. Manual. Moscow: Higher School of Economics National Research University, 2012.

[29] H. Brandt, Statistical methods of the analysis of observations. Moscow: Book on demand, 2012.

[30] Volkova P. A., Shipunov A. B. Statistical data processing in educational and research works. – M.: Forum, 2012. – 96 p.

[31] N. B. Paklin, V. I. Oreshkov, Business analytics. From data to knowledge, 2nd prod. SPb.: St. Petersburg, 2013.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731